

تحلیل سؤالات جامع آزمون نهایی پرستاری در دانشکده علوم پزشکی شاهرود

علی دادگری^{*}، حسین ابراهیمی^{**}، محمدرضا حسینی^{***}، یدالله مشایخی^{*}

چکیده:

هر آزمونی به عنوان ابزار سنجش و اندازه گیری می باید دارای روایی و پایایی کافی برای سنجش صفت مورد نظر باشد. با استفاده از آزمونهای ریاضی و آمار می توان آزمونها را تحلیل نمود و صحت عمل آنها را سنجید. هدف از انجام این تحقیق تحلیل سؤالات جامع آزمون نهایی پرستاری به عنوان اصلی ترین روش ارزشیابی تکوینی در دانشکده علوم پزشکی شاهرود می باشد. در این پژوهش هر یک از ۳۶۰ سؤال مجموع سؤالات آزمون جامع نهایی از نظر ضریب دشواری، ضریب تمیز، روایی، پایایی، وجود گزینه های انحرافی و غیره مورد ارزیابی و تحلیل قرار گرفتند تا نقاط ضعف و قوت هر سؤال و کل آزمون تعیین گردد. نتایج کسب شده در خصوص میزان دشواری هر سؤال در هر یک از خرده آزمونها نشان داد که دروس بهداشت جامعه و بهداشت مادران و نوزادان به ترتیب با ۴۲ و ۴۰ سؤال با ضریب دشواری بسیار بالا (۷۱٪-۱) دارای بیشترین سؤال آسان در مجموعه آزمونها بود. در حالی که در درس داخلی جراحی ۲ یا ۱۸ سؤال با ضریب دشواری بسیار بالا، کمترین سؤالات آسان را دارا بود. عمده ترین مشکلات طراحی سؤال در آزمون جامع نهایی پرستاری در دانشکده علوم پزشکی شاهرود طراحی سؤالات با ضریب سهولت بالا (دشواری اندک) و دارای گزینه های فاقد کارایی برای متمایز ساختن دانشجویان قوی از ضعیف می باشد علاوه بر این استفاده از آزمونهای مناسب برای تحلیل سؤالات آزمون اجتناب ناپذیر است. پژوهشگران استفاده از روش دو نیمه کردن را به جای روش کودر - ریچاردسون ۲۰ توصیه می نمایند.

واژه های کلیدی: ارزشیابی، آزمون، پرستاری.

مقدمه:

رایج ترین شکل ارزشیابی، ارزشیابی دانشجویان است (۲۰۱۱). آزمونهای مورد استفاده در مؤسسات آموزش عالی در کشور ما عمدتاً به صورت چهار گزینه ای طراحی می گردند. به ویژه هنگامی که آگاهی دانشجویان در حیطه شناختی موضوع مورد ارزشیابی می باشد (۳). هر

ارزشیابی یکی از بنیادی ترین وظایف هر مؤسسه آموزشی محسوب می گردد. ارزشیابی به اشکال و با اهداف مختلفی صورت می پذیرد. ممکن است ارزشیابی در مورد آزمون، مؤسسه، دانشجو و هر فرآیند اجرایی دیگر صورت پذیرد. در مؤسسات آموزشی

* عضو هیأت علمی گروه پرستاری - دانشگاه علوم پزشکی شاهرود - دانشکده پرستاری - تلفن: ۰۲۷۳-۳۳۳۶۰۱۷ (مؤلف مسئول).

** عضو هیأت علمی گروه پرستاری - دانشگاه علوم پزشکی شاهرود *** کارشناس ارشد انگل شناسی و عضو هیأت علمی معاونت تحقیقات و فناوری

* عضو هیأت علمی گروه بافت شناسی - دانشگاه علوم پزشکی شاهرود.

نظر گرفته شد.

مراحل تحلیل آزمون:

پاسخهایی که شرکت کنندگان در آزمون به سؤالات داده‌اند به عنوان داده‌ها مورد استفاده قرار می‌گیرند. ابتدا باید دانشجویان را بر اساس نمرات کسب شده ایشان در دو گروه قوی و ضعیف طبقه بندی کرد. برای طبقه بندی دانشجویان بر اساس تعداد، در آزمونه‌ای تا ۴۰ نفر شرکت کننده، ۱۰ نفر از احراز کنندگان با بهترین نمره آزمون و ۱۰ نفر از احراز کنندگان با پایین ترین نمرات آزمون را انتخاب می‌کنیم و در دو گروه قوی و ضعیف تقسیم می‌نمائیم. اگر تعداد آزمون شوندگان بیشتر از ۴۰ نفر باشند توصیه شده که ۲۷ درصد را به عنوان گروه قوی و ۲۷ درصد شرکت کنندگان را به عنوان گروه ضعیف در نظر بگیریم (۱). در تحلیل هر سؤال باید روی کارت مشخصات مربوط به آزمون و سؤال مربوطه به همراه گزینه‌ها نوشته شود و گزینه صحیح نیز تعیین گردد (تصویر شماره ۱).

موضوع آزمون:

داخلی جراحی ۱ درس: پرستاری اعصاب
سؤال: کدامیک از مایعات تزریقی زیر برای کاهش فشار مغزی بیمار مبتلا به ضربه سر مناسب تر است؟
الف - نرمال سالین ۰/۹٪ ب - رینگر لاکتات
ج - رینگر ساده د - مانیتول ۲۰٪

تصویر شماره ۱: کارت مشخصات آزمون

در پشت همین کارت می‌باید تعداد پاسخهای آزمون شوندگان به هر یک از گزینه‌های الف، ب، ج یا د به تفکیک گروه قوی و ضعیف مشخص گردد (تصویر شماره ۲).

آزمونی به عنوان ابزار سنجش و اندازه گیری می‌باید که دارای روایی و پایایی کافی برای سنجش صفت مورد نظر باشد. متأسفانه بسیاری از آزمونه‌های معلم ساخته و استاندارد نشده با اشکالات متعددی همراه هستند (۵،۴). با استفاده از آزمونه‌های ریاضی و آمار می‌توان صحت عمل آزمونها را سنجید و هر آزمونی را به استاندارد طراحی و اجرا نمود.

در این دانشکده دانشجویان در پایان دوره چهارساله پرستاری و پیش از فراغت از تحصیل و کسب مدرک کارشناسی در آزمون جامع نهایی شرکت می‌کنند. این آزمون از ۶ جزء، یا خرده آزمون تشکیل یافته است. خرده آزمونها شامل درس بیماریها و پرستاری داخلی - جراحی ۱، بیماریها و پرستاری داخلی جراحی ۲، پرستاری بیماریهای کودکان، پرستاری بیماریهای روانی، پرستاری بهداشت جامعه و پرستاری بهداشت مادران و نوزادان می‌باشد. هر خرده آزمونی از ۶۰ سؤال تشکیل یافته و در ۶ روز متوالی کل آزمونها برگزار می‌گردند.

هدف از انجام این تحقیق تحلیل آزمون جامع نهایی به عنوان اصلی ترین روش ارزشیابی از نظر ضریب دشواری، ضریب تمیز، پایایی و روایی در دانشکده علوم پزشکی شاهرود می‌باشد.

مواد و روشها:

در این پژوهش هر یک از ۳۶۰ سؤال آزمون جامع نهایی یک به یک مورد ارزیابی و تحلیل قرار گرفتند تا نقاط ضعف و قوت هر سؤال و کل آزمون تعیین گردد. جهت تحلیل سؤالات آزمون جامع نهایی در دانشکده علوم پزشکی شاهرود ۱۰ نفر از دانشجویان حائز بالاترین نمرات در هر خرده آزمون و ۱۰ نفر از دانشجویان حائز پایین ترین نمرات در هر خرده آزمون به ترتیب به عنوان گروههای قوی و ضعیف در نظر گرفته شدند. جهت تحلیل هر آزمون مراحل زیر در

نامید. با وجود این کاربرد واژه ضریب دشواری رایج تر است (۱).

تفسیر ضریب دشواری:

صاحب نظران معتقدند که ضریب دشواری بین ۰/۳ تا ۰/۷ بیشترین اطلاعات را در مورد میزان سهولت یا دشواری آزمون به ما می دهد. آنها توصیه می کنند در آزمونهای متداول درسی این محدوده به عنوان مناسب ترین محدوده در نظر گرفته شود، اما در برخی آزمونهای خاص با اهداف ویژه می توان ضریب دشواری بیشتر یا کمتر را مناسب تر دانست. مثلاً در آزمون انتخاب دانش آموزان برای پذیرش در دوره پیش دبستانی ضریب دشواری بسیار بالا توصیه می گردد. در حالی که برای پذیرش دانشجو در مقطع دکترای تریدید ضریب دشواری بسیار اندک یعنی سؤالات بسیار دشوار کارایی بیشتری دارند (۲).

در این پژوهش، گروههای قوی و ضعیف به تعداد ده نفر در هر گروه بر اساس نمره نهایی به دست آمده انتخاب شدند. سؤالات هر یک از خرده آزمونها یک به یک در گروههای قوی و ضعیف بررسی گردید و در کارت مربوط به هر سؤال تعداد پاسخهای داده شده به هر گزینه در هر دو گروه مشخص شد.

محاسبه ضریب تمیز:

طبق تعریف، ضریب تمیز و قدرت هر سؤال برای متمایز ساختن دانشجوی قوی از دانشجوی ضعیف است و با حرف d نشان داده می شود. برای تعیین ضریب تمیز از فرمول زیر استفاده شد:

$$\text{انتخابهای درست گروه پایین} - \text{انتخابهای درست گروه بالا} \\ \text{تعداد افراد در یک گروه}$$

بر اساس این فرمول مشخص می شود که هر چه ضریب تمیز سؤالی بیشتر باشد قدرت تفکیک دانشجوی قوی و ضعیف در این سؤال بیشتر است (۱).

موضوع آزمون:						
داخلی جراحی ۱						
درس: پرستاری اعصاب						
گروه	الف	ب	ح	د	بدون پاسخ	تعداد کل
قوی	۶	۲	۱۰	۰	۰	۲۰
ضعیف	۱۰	۸	۳	۰	۰	۲۰

تصویر شماره ۲: تعداد پاسخهای آزمون شوندگان به هر یک از گزینه های الف، ب، ج یا د به تفکیک گروه قوی و ضعیف

در تصویر شماره ۲ علاوه بر تعیین تعداد پاسخهای صحیح در دو گروه ضعیف و قوی، همچنین مشخص می گردد که آزمون شوندگان به کدام گزینه ها، چه صحیح و چه غلط بیشتر توجه کرده اند. به عبارت دیگر تعیین می گردد که:

(الف) در گروه قوی چه تعداد از شرکت کنندگان گزینه صحیح یا گزینه های غلط را برگزیده اند.

(ب) در گروه ضعیف چه تعداد از شرکت کنندگان گزینه صحیح یا گزینه های غلط را برگزیده اند.

بر اساس اطلاعات به دست آمده از بند الف و ب فوق می توان ضریب دشواری (ضریب سهولت) و ضریب تمیز هر سؤال را مشخص کرد (۱).

محاسبه ضریب دشواری:

طبق تعریف، دشواری درصد پاسخهای صحیح داده شده به هر سؤال است که با حرف P نشان داده می شود و با فرمول زیر محاسبه می گردد:

$$\frac{100 \times (\text{انتخابهای درست گروه پایین} + \text{انتخابهای درست گروه بالا})}{\text{تعداد افراد در یک گروه} \times 2}$$

محدوده ضریب دشواری بین صفر تا ۱۰۰ درصد متغیر است. بر اساس این فرمول مشخص می شود که هر چه ضریب دشواری سؤالی بزرگتر باشد آن سؤال آسانتر است و از این رو شاید بهتر باشد آن را ضریب سهولت

تفسیر ضریب تمیز:

ضرب تمیز بر خلاف ضریب دشواری که فقط از صفر تا صد در صد را شامل می‌شود، می‌تواند محدوده وسیع‌تری را شامل شود و حتی نمرات منفی را نیز احراز کند. اگر ضریب تمیز آزمونی ۰/۹ محاسبه شود نشان دهنده بالا بودن قدرت تمیز آن سؤال در آزمون است در حالی که اگر ضریب تمیز در سؤالی ۰/۲ شود بیانگر ضعف آن سؤال در تفکیک افراد قوی و ضعیف خواهد بود. چنانچه ضریب تمیز سؤالی منفی باشد، بی شک ساختار سؤال مربوطه دارای اشکالات اساسی است به نحوی که باعث شده دانشجوی قوی آن سؤال را غلط پاسخ دهد و دانشجوی ضعیف به آن سؤال پاسخ صحیح دهد. چنین سؤالی باید اصلاح و یا کلاً حذف گردد (۱).

تحلیل گزینه‌ها:

در هر آزمون چهار گزینه‌ای، تنها یک گزینه کاملاً صحیح و سه گزینه غلط وجود دارد. هدف از کاربرد این سه گزینه سنجش قدرت دانشجو در انتخاب گزینه صحیح در میان چهار گزینه پیشنهادی است. گزینه‌های غلط نباید نقش راهنمایی کننده داشته باشد. تحلیل گزینه‌ها به طراح سؤال این امکان را می‌دهد تا نکات انحرافی یا راهنمایی کننده هر گزینه که از نظر طراح سؤال دور مانده را شناسایی و به اصلاح ساختار آزمون بپردازد (۲، ۱). در تحلیل گزینه‌ها باید دید که چند نفر از دانشجویان گروه ضعیف گزینه ناصحیح را انتخاب کرده یا انتخاب نکرده‌اند. اگر هیچ یک از دانشجویان گروه ضعیف گزینه ناصحیحی را انتخاب نکنند، احتمالاً آن گزینه آن چنان غلط آشکاری دارد که توانسته به شکلی آزمون شونده را به ناصحیح بودن آن هدایت کند و یا گزینه صحیح آن چنان هدایت کننده و آشکار طراحی شده است که سایر گزینه‌ها اصلاً توجه آزمون شونده را به خود جلب نکرده‌اند. چنین حالتی در گروه دانشجویان ضعیف بررسی می‌شود و در تعداد نمونه‌های بالا از نظر

آزمون شونده کارایی بیشتری دارد (۱).

پایایی آزمون:

در سؤالات چهار گزینه‌ای همواره ارزیابی پایایی آزمون مورد نظر و توجه متخصصین ارزشیابی دوره است (۴، ۲). بررسی پایایی آزمون میزان صحت آزمون برای سنجش صفت مورد نظر را به ما نشان می‌دهد. برای تعیین میزان پایایی (ضریب همبستگی) آزمونها روشهای مختلفی به کار گرفته می‌شوند. دو روشی که بیشتر کاربرد دارند، به نام "روش دو نیمه کردن آزمون" و "روش کودر - ریچاردسن ۲۰" می‌باشند.

الف) روش دو نیمه کردن:

$$r_{11} = \frac{2r \frac{1}{2} \frac{1}{2}}{1 + r \frac{1}{2} \frac{1}{2}}$$

$$r_{11} = \text{reliability}$$

$$r \frac{1}{2} \frac{1}{2} = \text{Correlation of two halves}$$

در این روش هر آزمون بر اساس سؤالات فرد و زوج، به صورت دو آزمون در نظر گرفته می‌شود و ضریب همبستگی این دو آزمون با فرمول اسپیرمن - براون ارزیابی می‌گردد و به عنوان پایایی آزمون در نظر گرفته می‌شود (۲).

ب) روش کودر - ریچاردسن ۲۰

$$r_{11} = \left[\frac{n}{n-1} \right] \left[1 - \frac{\sum pq}{O^2} \right]$$

$$r_{11} = \text{reliability of the exam}$$

$$n = \text{number of questions}$$

$$O^2 = \text{variance}$$

$$p = \text{ratio of correct answers}$$

$$q = 1 - p$$

$$\sum pq = \text{sum of } pq$$

نتایج:

نتایج حاصل از تحلیل سؤالات آزمون جامع نهایی، میزان ضریب دشواری هر سؤال در هر یک از خرده آزمونها نشان داد که دروس پرستاری بهداشت جامعه،

جدول شماره ۱: ضریب دشواری سؤالات به تفکیک خرده آزمونها

خرده آزمونه‌های پرستاری	۰-۰/۴ مشکل	۰/۷-۰/۴۱ متوسط	۰/۷۱-۱ آسان
کودکان	۷	۱۳	۴۰
بهداشت جامعه	۷	۱۱	۴۲
بهداشت مادران	۸	۱۲	۴۰
داخلی جراحی ۱	۱۰	۱۹	۳۱
داخلی جراحی ۲	۱۴	۲۸	۱۸
روانی	۹	۱۴	۳۷

سؤالات هر خرده آزمون بر اساس ضریب دشواری به سه گروه آسان (با ضریب دشواری ۱-۰/۷)، متوسط (با ضریب دشواری ۰/۷-۰/۴۱) و مشکل (با ضریب دشواری ۰/۴-۰/۷۱) تقسیم شده‌اند.

جدول شماره ۲: ضریب تمیز سؤالات به تفکیک خرده آزمونها

خرده آزمونه‌های پرستاری	<۰ منفی	۰-۰/۴ اندک	۰/۴۱-۱ بالا
کودکان	۱۳	۴۵	۲
بهداشت جامعه	۸	۵۲	۰
بهداشت مادران	۱۲	۴۶	۲
داخلی جراحی ۱	۸	۴۵	۷
داخلی جراحی ۲	۱۰	۴۶	۴
روانی	۸	۴۸	۳

سؤالات هر خرده آزمون بر اساس ضریب تمیز به سه گروه با ضریب تمیز منفی (کمتر از صفر)، اندک یا ضریب تمیز (۰/۴-۰) و بالا یا ضریب تمیز (۱-۰/۴۱) تقسیم شده‌اند.

یکی دیگر از اهداف اختصاصی این تحقیق بررسی گزینه‌های هر سؤال از نظر جلب توجه آزمون‌شوندگان بود. بررسی نتایج مؤید این واقعیت بود که در برخی سؤالات یک، دو و یا حتی هر سه گزینه ناصحیح فاقد کارایی لازم بودند. نتایج نشان داد که در کل آزمون، ۳۲ سؤال به گونه‌ای طراحی شده بودند که حتی در گروه دانشجویان ضعیف نیز هیچ فردی گزینه‌های ناصحیح را انتخاب نکرده بود. در بین خرده آزمونها نیز دروس پرستاری بیماریهای روانی و پرستاری بیماریهای داخلی جراحی ۲، به ترتیب با ۱۴ و صفر سؤال دارای ۳ گزینه فاقد کارایی، بیشترین و کمترین آمار در این زمینه را دارا بودند (جدول شماره ۳).

نکته مهمی که از نظر پژوهشگران نیاز به توجه بیشتری دارد ناکارآمد بودن روش "کودر - ریچاردسن" ۲۰ در تعیین پایایی آزمونها چهارگزینه‌ای است، چرا که این روش مختص آزمونها صحیح - غلط می‌باشد و نباید از آن در سؤالات چهارگزینه‌ای استفاده کرد (۲) با وجود این پژوهشگران با استفاده از این روش سعی

پرستاری بیماریهای کودکان و پرستاری بهداشت مادران و نوزادان به ترتیب با ۴۰، ۴۰ و ۴۰ سؤال با ضریب دشواری بسیار بالا (۰/۷۱-۱) دارای بیشترین سؤال آسان در مجموعه آزمونها بودند. در حالی که درس داخلی جراحی ۲ با ۱۸ سؤال با ضریب دشواری بسیار بالا کمترین سؤالات آسان را دارا بود (جدول شماره ۱). با استفاده از فرمول ضریب تمیز در مورد هر سؤال آزمون، دیده شد که می‌توان سه دسته ضریب تمیز را مشخص نمود. ضریب تمیز کمتر از صفر (یا منفی)، ضریب تمیز بین صفر تا ۰/۴ و ضریب تمیز بین ۰/۴۱ تا ۱. بر اساس نتایج به دست آمده از ضریب تمیز هر یک از سؤالات در خرده آزمونها دیده شد که در دروس پرستاری بیماریهای کودکان و پرستاری بهداشت مادران - نوزادان، با دارا بودن ۱۳ و ۱۲ سؤال با ضریب تمیز منفی بیشترین تعداد سؤال را در این گروه داشتند. در حالی که کمترین تعداد سؤال با ضریب تمیز منفی (۸ سؤال) در درس داخلی جراحی ۱، پرستاری بهداشت جامعه و پرستاری روانی دیده شد (جدول شماره ۲).

جدول شماره ۳: گزینه‌های فاقد کارایی در سؤالات به تفکیک خرده آزمونها

خرده آزمونها	گزینه ۰	گزینه ۱	گزینه ۲	گزینه ۳
کودکان	۱۷	۱۹	۲۳	۱
بهداشت جامعه	۱۵	۲۵	۱۶	۴
بهداشت درمان	۱۷	۱۶	۱۶	۱۱
داخلی جراحی ۱	۱۶	۲۶	۱۶	۲
داخلی جراحی ۲	۳۶	۱۹	۵	۵
روانی	۳۲	۱۰	۱۲	۱۴

گزینه‌های فاقد کارایی به گزینه‌های اطلاق می‌گردد که به شکل واضحی ناصحیح بودن آنها مشخص است، به گونه‌ای که در گروه دانشجویان ضعیف نیز حتی یک نفر آنها را به عنوان پاسخ انتخاب نکرده است.

داشتند اختلاف موجود را به شکل ملموس‌تری نشان دهند. لذا کاربرد روش دو نیمه کردن آزمون، نتایج قابل اعتمادتری نسبت به روش کودر - ریچاردسن ۲۰ به ما نشان می‌دهد (جدول شماره ۴).

جدول شماره ۴: ضریب همبستگی یا پایایی خرده آزمونها به دو روش "کودر-ریچاردسن ۲۰" و "دو نیمه کردن آزمون"

خرده آزمونها	روش A	روش B
کودکان	۰/۷۹	۰/۲۶
بهداشت جامعه	۰/۸۶	۰/۶۸
بهداشت مادران	۰/۹۰	۰/۶۴
داخلی جراحی ۱	۰/۸۳	۰/۷۷
داخلی جراحی ۲	۰/۷۳	۰/۶۴
روانی	۰/۹۳	۰/۲۵

مقایسه ضریب همبستگی آزمون به دو روش کودر-ریچاردسن ۲۰ و دو نیمه کردن آزمون نشان دهنده اختلاف نتایج به دست آمده می‌باشد. A روش کودر-ریچاردسن ۲۰، B روش دو نیمه کردن آزمون.

بحث:

سؤالات چهار گزینه‌ای شکل متداول آزمون برای ارزشیابی دانشجویان است (۴،۳،۲). اشکالات و نقایص طراحی سؤالات چهار گزینه‌ای در تحقیقات مختلف بررسی شده‌اند. اشکالات ساختاری در طراحی این سؤالات بکرات دیده می‌شود. از شایع‌ترین اشکالات ساختاری، نارسا بودن ساقه سؤال، مبهم بودن سؤال یا گزینه‌ها، کاربرد منفی در منفی در سؤال، عدم تطابق گزینه‌ها با ساقه سؤال از نظر جمله بندی و منطقی و غیره باشد (۴).

آن چنان که از نتایج این تحقیق نیز بر می‌آید عمده‌ترین مشکلات طراحی سؤال در آزمون جامع نهایی پرستاری در دانشکده علوم پزشکی شاهرود طراحی سؤالات با ضریب سهولت بالا (دشواری اندک) و دارای گزینه‌های فاقد کارایی برای متمایز ساختن دانشجویان قوی از ضعیف می‌باشد. با وجود این، پایایی خرده آزمونها در بیشتر موارد خوب و قابل قبول بود. با توجه به متداول بودن چنین اشکالاتی در تمامی مؤسسات آموزشی، استاندارد کردن سؤالات در آزمونها معلم ساخته در علوم پزشکی ضرورتی اساسی در آموزش می‌باشد (۵،۳). چنین سؤالاتی را می‌توان برای استفاده در آزمونها بعدی در بانک سؤالات دانشکده ذخیره نمود.

بدون استاندارد کردن آزمون، قضاوت در مورد آگاهی و دانش فراگیران بسیار دشوار و گاه غیر ممکن می‌نماید و بدیهی است که چنین سؤالاتی نمی‌تواند اعتبار لازم برای سنجش آموخته‌های دانشجویان را داشته باشند (۴،۲،۱).

پژوهشگران پیشنهاد می‌کنند که دانشگاه و مؤسسات آموزش عالی سؤالات تمامی آزمونها را با روشهای فوق استاندارد نمایند.

تشکر و قدردانی:

بدینوسیله از کلیه کسانی که ما را در این امر یاری نمودند

قدردانی می‌گردد.

منابع:

- ۱- سیف علی اکبر. اجراء و نمره در: سیف علی اکبر. روشهای سنجش و اندازه گیری. انتشارات دوران. تهران: چاپ چهارم، ۱۳۷۸، ۲۷۱-۳۰۵.
- ۲- سیف علی اکبر. پایایی آزمون، در: سیف علی اکبر. روشهای سنجش و اندازه گیری. انتشارات دوران، تهران: چاپ چهارم، ۱۳۷۸، ۳۷۷-۲۳۹.
- ۳- نوریان جواد؛ دادگری علی. مقایسه روشهای یادگیری و ارزشیابی. پژوهش در علوم پزشکی: ویژه نامه همایش آموزش پزشکی. اصفهان: ۱۳۷۵، ۱۰۸-۹.
- 4- Gareth H.; Elzubier M. Imprecise terms in UK medical multiple - choice questions: what examiners think they mean. Med Educ, Glasgow: UK, 32: 343-50, 1998.
- 5- McLeod PJ.; Snell L. Student-generated MCQs. Med Educ, Glasgow: UK, 18: 23-4, 1996.
- 6- Polit DF.; Hungler BP. Data Quality. In: Polit DF.; Hungler BP. Essentials of nursing research: methods appraisal and utilization: From Lippincotte Company. NewYork: USA, Chapt 9. 3rd ed. 249, 1993.